Prognostic models for mortality after cardiac surgery in patients with infective endocarditis: a systematic review and aggregation of prediction models

Author list:

Borja M. Fernandez-Felix[1,2], Laura Varela Barca[3], Esther Garcia-Esquinas[2,4,5], Andrea Correa-Pérez[1,6], Nuria Fernández-Hidalgo[7,8], Alfonso Muriel[1,2], Jesus Lopez-Alcalde[1,2,6,9], Noelia Álvarez-Diaz[10], Jose I. Pijoan[2,11], Aida Ribera[2,12], Enrique Navas Elorza[13], Patricia Muñoz[14], Mª Carmen Fariñas[15], M. Ángel Goenaga[16], Javier Zamora[1,2,17]

Affiliations:

[1] Clinical Biostatistics Unit, Hospital Universitario Ramon y Cajal, Madrid, Spain
[2] CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain
[3] Department of Cardiovascular Surgery Fundacion Jimenez Diaz University Hospital, Madrid, Spain
[4] Department of Preventive Medicine and Public Health. School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain
[5] IdiPaz (Hospital Universitario La Paz-Universidad Autónoma de Madrid), Madrid, Spain
[6] Faculty of Medicine. Universidad Francisco de Vitoria, Madrid, Spain
[7] Servei de Malalties Infeccioses, Hospital Universitari Vall d'Hebron, Barcelona, Spain.
[8] Red Española de Investigación en Patología Infecciosa (REIPI), Instituto de Salud Carlos III, Madrid, Spain
[9] Institute for Complementary and Integrative Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland
[10] Medical Library, Hospital Universitario Ramon y Cajal, Madrid, Madrid, Spain
[11] Hospital Universitario Cruces/OSI EEC; Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain
[12] Cardiovascular Epidemiology and Research Unit, Hospital Universitari Vall d'Hebron, Barcelona, Spain
[13] Department of Infectology, Hospital Universitario Ramon y Cajal, Madrid, Spain
[14] Clinical Microbiology and Infectious Diseases Service, Hospital General Universitario Gregorio Marañón, Madrid. Instituto de Investigación Sanitaria Gregorio Marañón. CIBER Enfermedades Respiratorias-CIBERES. Facultad de Medicina, Universidad Complutense de Madrid, Spain
[15] Infectious Diseases Service. Hospital Universitario Marqués de Valdecilla-IDIVAL. Universidad de Cantabria, Santander, Spain.
[16] Infectious Diseases Service. Hospital Universitario Donostia. IIS Biodonostia. OSI Donostialdea. San Sebastián, Spain
[17] WHO Collaborating Centre for Global Women's Health, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

41   Corresponding author:

42   Complete name: Borja Manuel Fernández Félix

43   Email address: borjamanuel86@gmail.com

44   ORCID: http://orcid.org/0000-0002-8798-019X

45   Postal address: Hospital Universitario Ramón y Cajal. Ctra. de Colmenar Viejo, Km. 9,100

46   (28034) MADRID

47   Phone number: +34 91 3368103

48   Fax number: +34 91 3369016

49   Category: Systematic review

Abstract

Background: There are several prognostic models to estimate the risk of mortality after surgery for active infective endocarditis (IE). However, these models incorporate different predictors and their performance is uncertain.

Objective: We systematically reviewed and critically appraised all available prediction models of post-operative mortality in patients undergoing surgery for IE, and aggregated them into a meta-model.

Data sources: We searched Medline and EMBASE databases from inception to June 2020.

Study eligibility criteria: We included studies that developed or updated a prognostic model of post-operative mortality in patient with IE.

Methods: We assessed the risk of bias of the models using PROBAST (Prediction model Risk Of Bias ASsessment Tool) and we aggregated them into an aggregate meta-model based on stacked regressions and optimized it for a nationwide registry of IE patients. The meta-model performance was assessed using bootstrap validation methods and adjusted for optimism.

Results: We identified 11 prognostic models for post-operative mortality. Eight models had a high risk of bias. The meta-model included weighted predictors from the remaining three models (*i.e.*, EndoSCORE, specific ES-I and specific ES-II), which were not rated as high risk of bias and provided full model equation. Additionally, two variables (*i.e.*, age and infectious agent) which had been modelized differently across studies, were estimated based on the nationwide registry. The performance of the meta-model was better than the original three models, with the corresponding performance measures: C-statistics 0.79 (95% CI 0.76 to 0.82), calibration slope 0.98 (95% CI 0.86 to 1.13) and calibration-in-the-large -0.05 (95% CI -0.20 to 0.11).

Conclusions: The meta-model outperformed published models and showed a robust predictive capacity for predicting the individualized risk of post-operative mortality in patients with IE.

3

75   *Protocol Registration:* PROSPERO (registration number CRD42020192602)

76   *Key words:* Prognostic models, systematic review, meta-model, aggregation, validation,

77   infective endocarditis.

78

Background

Infective endocarditis (IE) is an uncommon but severe disease with a high mortality rate. Its current estimated incidence is 3-10 episodes per 100.000 person-years, while its in-hospital mortality rate ranges between 15% and 40% (1,2). Management of IE is often complex and, the decision whether to perform surgery remains a challenge because of the high mortality rate associated with the procedure. For that reason, it is estimated than less than half of the patients with surgical indication finally undergo cardiac surgery (3); which leads to a significantly decreased chance of survival (4). In this context, there has been a great interest in modeling prognosis of patients with IE to accurately estimate the risk of mortality in patients undergoing surgery for IE, and to help in the decision-making processes.

Prognostic models are mathematical equations that relates multiple variables for a particular individual to the probability of post-operative mortality. In the last decade, several IE prognostic models using preoperative patient´s-related and IE-specific factors, have been proposed. Unfortunately, these models have not been implemented in guidelines or are rarely applied in clinical practice. The poor adoption of these models could be a consequence of a shared perception of their limited validity because they have usually been built in relatively small cohorts and lack of external validation. Consequently, researchers carry on developing new models using their own data without considering prior knowledge, which leads to a scenario with multiple prognostic models of dubious validity. Therefore, we aimed to systematically review and critically appraise all available prediction models for post-operative mortality after cardiac surgery in patients with IE. We also aimed to aggregate those models with low risk of bias into a meta-model based on stacked regressions.

Methods

The protocol for this study was registered on PROSPERO (registration number CRD42020192602). We designed this systematic review according to the recent guidance(5,6), and reported its results following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (7) and TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) (8,9) recommendations.

Literature search

We searched Medline through Ovid and Embase through Elsevier from inception to 01/06/2020. We conducted a literature search to identify all potential studies for inclusion, without any language or publication dates restriction. We used the methodologic filter developed by Geersing et al. for prediction models research in MEDLINE (10), which was adapted for EMBASE. We added terms related to cardiac surgery and endocarditis. We further searched bibliographic references of included articles to identify other potential eligible studies. Complete search strings are shown in **Supplementary Material: S1.**

Eligibility criteria

We included original studies that developed prognostic models, with or without external validation, to predict the risk of post-operative mortality after cardiac surgery in patients with IE, as well as studies that updated previously published models. We accepted the authors` definition of post-operative mortality (either 30 days and/or in-hospital mortality), but excluded models that predicted mortality as part of a composite adverse outcome. Titles, abstracts, and full texts were screened for eligibility in pairs by three reviewers independently (BMFF, LVB, ACP) using EPPI-Reviewer 4 (11). Discrepancies were resolved by consensus.

Data extraction

Data extraction of included articles was done by three reviewers independently (pairs from BMFF, LVB, ACP). Discrepancies were solved by consensus. Reviewers used a standardized data extraction form based on CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) (6). We extracted data on the following items: general information of the study, source of data, participants´ characteristics, outcome definition and time of occurrence, candidate predictors, and analysis methods. (**Supplementary Material: S2**). When the completed model equation or relevant data were not provided, we contacted the correspondence authors to require this information.

Risk of bias assessment

We used a standardized form based on PROBAST (PRediction model risk of Bias ASsessment Tool) (12,13) to evaluate risk of bias (RoB) and applicability. We used the PROBAST definition of RoB. Concerns regarding the applicability of a primary study would arise when the population, predictors, or outcomes of the study differed from those specified in our review question. RoB and applicability were assessed by two independent reviewers (pairs from BMFF, LVB, ACP). We evaluated the relevant items on the following domains: Participants, predictors, outcome and analysis. Each domain was rated as a *high*, *low* or *unclear* RoB, and as providing *high*, *low* or *unclear* concerns regarding applicability. Any discrepancies were discussed between reviewers and resolved through discussion. The supplementary material provides details on critical appraisal and applicability (**Supplementary Material: S3**).

GAMES registry

We used the nationwide GAMES – Grupo de Apoyo al Manejo de la Endocarditis infecciosa en España – (14) registry as the validation dataset, to estimate existing models` weights for the meta-model development and its validation, and to externally validate the previously

7

published models. Since January 2008, all consecutive episodes of IE in 34 Spanish hospitals

were prospectively registered in GAMES using a standardized form. Regional and local ethics

committees approved the study, and patients gave their informed consent in each center. For

the present study, we selected all the infective episodes (n=1,453) registered in the GAMES

cohort involving adult patients (aged ≥18 years) who had undergone cardiac surgery with

preoperative diagnosis of active IE. From these, 354 (24.4%) died after surgery (273 in the

first 30 days and the remaining 81 during hospitalization). Assessment of predictors was done

in an unblinded manner (i.e. with knowledge of the participant's outcome). **Supplementary**

Material: Table S1 shows the main descriptive characteristic of patients in the validation

nationwide registry.

Statistical analyses

Model aggregation was based on stacked regressions (15). This methodology allows the

synthesis of models collated in a systematic review into a meta-model using a validation

dataset (16,17). We did not consider for aggregation the models that did not report the full

equation or the models that were classified as high risk of bias. Stacked regressions used the

linear predictor of each model as a co-variable in the meta-model, to subsequently created a

linear combination of model predictions. That is, the original coefficients of each model are

weighted by an independent parameter estimated in the meta-model, so that the models with

worse performance in the validation dataset are penalized more. When aggregation of the

coefficients was not possible, either because the definition of the predictor from primary

studies was too heterogeneous or because predictors had been modeled differently in the

published models (for instance, a numerical variable treated as a continuous predictor in one

model and being categorized at different cut-points in the others), these predictors were

dropped, and were included in the meta-model as independent covariables to re-estimate their

174   coefficients entirely from scratch based on the validation dataset. Non-linear relationships for

175   continuous predictors were tested using fractional polynomials (18).

176   Predictors with missing data in the validation dataset were imputed under the missing at

177   random assumption using multiple imputation with chained equations (19). We included all

178   predictors and the outcome in the imputation models to ensure compatibility.

179   (**Supplementary Material: S4**). Imputations checks were completed by looking at the

180   distributions of imputed values to ensure plausibility. We generated 10 multiple imputed

181   datasets and all primary analyses were performed in each imputed dataset. Pooled parameters

182   were estimated both in the aggregation and validation processes using Rubin's rules (20).

183   The meta-model validation was assessed in terms of discrimination (*i.e.*, through the use of

184   the C-statistic, with values from 1 indicating perfect discrimination to 0.5 no discrimination)

185   and calibration (*i.e.*, through the calibration slope and calibration-in-the-large [CITL], with 1

186   and 0 as ideal values, respectively; as well as with calibration plots). Calibration plots

187   represent the average predicted probability for risk groups categorized using deciles of

188   predicted probability against observed proportion in each group, and fitting a lowess smoother

189   to show calibration across the entire range of predicted probabilities at the individual-level

190   (21,22). For the calibration plots we used the average predicted probabilities for individuals

191   by pooling the imputed datasets using Rubin's rules (20). Because the meta-model was

192   optimized to the validation dataset, we assessed its optimism-corrected performance measures

193   by applying bootstrap validation with 500 replicates. As sensitivity analyses, we tested all

194   model performance regardless of their critical appraisal. In addition, the meta-model

195   performance was assessed only for 30-days mortality to investigate the meta-model

196   robustness. To facilitate the use of the model, an online version of the prognostic tool was

197   implemented in Evidencio (https://www.evidencio.com/). All analyses were performed using

198   Stata software version 16 (23).

199

Results

Search results and study selection

We retrieved 4,862 titles through our systematic search combining Medline and Embase. From these, 684 duplicate references were identified. Of 4,178 titles assessed by title and abstract, 34 studies were retained for full text screening, and 2 additional studies were detected in the bibliographic references of these articles. Nine studies describing 11 prediction models met the inclusion criteria (**Figure 1 and Supplementary Table S2**).

Source of data and participants

All included prognostic model development studies were published between 2011 and 2018. Six used data from a study cohort (three of them from a single center (24–26) and three from multiple centers (27–29)); two studies used data from multicenter registries (30,31); and one study used data from both a multicenter cohort and a local clinical registry (32). Eight studies used data from patients in Europe (Spain, Italy, France or Portugal) and one from patients in North America. Participants were recruited between 1980 and 2015. (**Supplementary Table S3**).

Outcomes

Three models were developed to predict any death occurring before discharge or within 30 days of surgery (24,26,30), five models to predict any death occurring before discharge (25,29,31,32), and the remaining three as death within 30 days of surgery (28, 29). The incidence of deaths varied between 8.2% and 29.2% (**Table 1**).

Predictors

The number of candidate predictors considered in the models ranged from 15 to 57 and included patient-, clinical-, surgery- and IE-related factors. The number of parameters

223 retained in the final models ranged from 2 to 15 (**Table 1**): The most common factors were

224 critical preoperative state (n=9), renal failure (n = 7), age (n = 6), New York Heart

225 Association (NYHA) classification of functional status (n=6), paravalvular complications (n =

226 6) and infection etiology (n = 5). The predictor definitions and the models' composition are

227 shown in the **Supplementary Table S4 and Table S5**.

228 *Model development and presentation*

229 Sample sizes for models' development varied between 128 and 13,617 patients, and the

230 number of events ranged from 21 to 1,117. Only two models from the same study adequately

231 informed the handling of missing data (28), and these used complete data analyses. Logistic

232 regression analysis was the most common modelling technique (n = 9), while logistic mixed

233 effects (27) and logistic Generalized Estimating Equation (GEE) models (30) were only used

234 in one model development each. Nine models used univariable analyses to select the

235 candidate predictors. In nine out of eleven models the number of events per parameter (EPP)

236 assessed for inclusion in the final model was lower than the minimum required for

237 development of a new prediction model, based on the sample size estimation proposed by

238 Riley et al.(33,34) (**Supplementary Table S6**). The method of predictors selection during

239 multivariable modelling was backward selection in three models (25,32), stepwise selection in

240 two models (29,31), and an automatic algorithm based on Akaike information criteria in

241 multiple bootstrap samples in the other two models, with predictors selected in at least 70% of

242 the bootstrapped samples being included in the final model (28). Four models did not inform

243 about the method used to select predictors. (**Table 1**)

244 In seven out of 11 models the authors omitted the complete model equation (in five of them

245 correspondence authors did not respond when were asked for further details)

246 (**Supplementary Table S7**). Nine models were presented as a scoring system, and two of

247 them included nomograms.

Model performance

The model performance was assessed in terms of discrimination through the C-statistic in all models. Nevertheless calibration was often wrongly assessed using the Hosmer-Lemeshow test (35) in six models. Only three models (26,28) used calibration slopes and CITL. Eight models were internally validated: three models were evaluated by bootstrapping with correction for optimism (27,28), one was assessed through the 0.632 bootstrap method (25), two used temporal split samples (32) and two used random split samples (29,30). Three models only estimated the apparent performance (24,26,31). Three models were externally validated in the same development study using very small sample sizes, with only 18 events in the Olmos' model (29) and 21 in the Gatti's models (32). Clinical utility of the models was never assessed.

Risk of bias

The RoB was high in eight models, unclear in one (27) and low in the remaining two (28) (**Table 1, Supplementary Table S8 and Figure S1**). Two of the eight models with high RoB scored at "high risk" in the participants domain. Eight models scored at "high risk" in the analysis domain. Most of the models had small sample sizes and even the number of EPP was close to 1 in several models, increasing the risk of overfitting (34). Many studies decided model predictors based on univariable analysis, three reported only the apparent performance and two used random splitting validation. The calibration was sub-optimally assessed in all models classified as high risk of bias, with most of them using the Hosmer-Lemeshow test.

Derivation of the Meta-model

The eight models with high RoB were excluded from the statistical synthesis so that only the EndoScore, Specifics EuroSCORE-I (Specific ES-I) and EuroSCORE-II (Specific ES-II) models were aggregated in the meta-model. The model developed by Di Mauro (EndoSCORE) (27) included 15 parameters, while the other two (Specific ES-I and Specific

273 ES-II) developed by Fernández-Hidalgo (28), presented 10 and 9 parameters respectively,

274 from the EuroSCORE models predictors (35, 36) and IE-related factors (**Table 2 and**

275 **Supplementary Table S7**). The dependent variable for the meta-model was mortality (either

276 30-days or in-hospital).

277 To construct the meta-model, we first calculated the linear predictors (LP) from EndoSCORE,

278 Specific ES-I and Specific ES-II for each observation in the validation dataset, after dropping

279 the parameters for age and infection etiology because these variables were modelized

280 differently in the different studies. Subsequently, we adjusted the meta-model using a logistic

281 regression model, which incorporated the LPs as co-variables, to estimate the models' weights

282 for aggregation, as well as the predictors for age (treated as continuous) and infection etiology

283 (categorized into three groups: *Staphylococcus spp.*, fungi and other microorganisms) to re-

284 estimate the coefficients from scratch. The meta-model included the predictors considered in

285 at least one of the three original models. These are patient-related factors (i.e. age, gender,

286 renal failure, prior cardiac surgery, chronic pulmonary disease, pulmonary hypertension and

287 left ventricular ejection fraction), clinical presentation-related factors (i.e. critical preoperative

288 state, New York Heart Association (NYHA) classification of functional status), surgery-

289 related factors (i.e. presence of paravalvular complications (abscess and/or fistulae), urgency

290 of procedure and number of treated valves/prostheses) and finally IE-related factors (i.e.

291 valve location and infection etiology) (**Supplementary Table S5**). The magnitude of the

292 associations of the predictive factors with mortality is shown in **Table 2**.

293 *Validation of the models*

294 The three prediction models considered for aggregation and the meta-model were validated in

295 the GAMES registry. The C-statistics and their 95% confidence intervals (95%CI) for the

296 published models were: 0.759 (95% CI 0.731 to 0.788) for EndoSCORE, 0.758 (95% CI

297 0.731 to 0.786) for Specific ES-I, and 0.762 (95% CI 0.735 to 0.789) for Specific ES-II. The

298   optimism adjusted C-statistic for the meta-model was 0.79 (95% CI 0.76 to 0.82) (**Figure 2**).

299   Calibration slopes were < 1 for all published models: 0.80 (95% CI 0.69 to 0.92) for

300   EndoScore, 0.82 (95% CI 0.70 to 0.94) for Specific ES-I, and 0.76 (95% CI 0.65 to 0.87) for

301   Specific ES-II. CITL was 0.58 (95% CI 0.44 to 0.71) for EndoSCORE and 0.62 (95% CI 0.48

302   to 0.76) for Specific ES-II, and -0.02 (95% CI -0.16 to 0.11) for Specific ES-I. Optimism

303   adjusted calibration measures for the meta-model were 0.98 (95% CI 0.86 to 1.13) for the

304   slope and -0.05 (95% CI -0.20 to 0.11) for CITL (**Figure 2**). The calibration plots for the

305   three previously published models and the meta-model are shown in **Figure 3**.

306   Sensitivity analysis showed that the meta-model had better overall performance than all

307   published models regardless of their quality assessment (**Supplementary Figure S2**).

308   Moreover, even though the meta-model was not fitted for the 30-days mortality outcome, it

309   outperformed the three models used for model aggregation. (**Supplementary Figure S3**)

310

Discussion

Summary of findings

311 In this systematic review of prediction models for post-operative mortality in patients with

314 infective endocarditis, we identified and critically appraised 11 models developed in 9 studies.

315 The predicted outcome varied between studies (in-hospital, 30-days or both in-hospital or 30-

316 days mortality). Of the eleven prognostic models, only two had low RoB and one unclear; the

317 remaining eight models had high RoB mainly owing to poor statistical methods used, which

318 suggests that their predictive performance when used in practice is probably lower than that

319 reported. The sample sizes used to develop the models were limited and this is a well-known

320 problem that leads to inaccurate predictions and consequently incorrect healthcare decisions

321 in practice (34).

322 Four out of the 11 published models reported the full model equation required for a models'

323 aggregation and a complete independent external validation as recommended by reporting

324 guidelines (8,9). Two models' equations were recovered after request to the corresponding

325 authors. Three models that were flagged as low or unclear RoB were aggregated to build the

326 meta-model.

327 Our meta-model included as predictors age, gender, renal failure, prior cardiac surgery,

328 chronic pulmonary disease, pulmonary hypertension, left ventricular ejection fraction, critical

329 preoperative state, New York Heart Association (NYHA) classification of functional status

330 presence of paravalvular complications (abscess and/or fistulae), urgency of procedure,

331 number of treated valves/prostheses, valve location and infection etiology. It showed better

332 performance than the original models. We investigated the internal validity of the meta-model

333 using bootstrap validation, and the results indicate there was no substantial over-optimism and

334 that the validation sample was sufficiently large to combine and update the published models.

335 Therefore, the meta-model is likely less prone to over-optimism and more generalizable to

336 new patient populations or settings, because it was built from the evidence of several patient

337 cohorts and optimized to a nationwide registry.

338 *Strengths and limitations*

339 To our knowledge, this is the first systematic review with specific focus on prediction models

340 of post-operative mortality in patients with infective endocarditis, with a thorough evaluation

341 of the RoB, and using an external validation cohort to build a meta-model. We only combined

342 the prediction models with low or unclear RoB and adjusted them to a new patient population.

343 We used multiple imputation of predictors to avoid loss of useful information. The resulting

344 meta-model incorporated prior knowledge optimally and outperformed previously published

345 models.

346 Our study has some limitations. The outcome definition in the validation dataset was either

347 30-days or in-hospital post-operative mortality, and the outcome definition in the three

348 models used for aggregation was 30-days mortality. Despite this difference a sensitivity

349 analysis showed that the meta-model outperformed all published models when we explored its

350 performance for the 30-days mortality. Two out of the three models considered for

351 aggregation were developed in the same cohort. This circumstance increases the probability

352 that the same predictors were included in both models and, therefore, it could magnify their

353 associations with the outcome in the meta-model. However, we think that the impact of this

354 magnification is limited because the weight of the ES-I model is relatively small compared to

355 the other two models. Unfortunately, although we identified 11 prediction models in our

356 systematic review, we were only able to validate the models for which the complete model

357 equation was available. We cannot rule out the presence of publication bias in our review.

358 Unpublished studies are likely to be of poor quality (small, overfitted, and with poor

359 predictive performance). Therefore, it is very likely that they would have been excluded from

360 our meta-model due to their high risk of bias. So the impact of this bias is expected to be low.

361   Although the definition of predictors in GAMES registry was standardized, these could differ

362   from definitions of published studies.

363   *Comparison to existing studies*

364   Most studies to develop new prediction models are based on small sample sizes and the

365   modelling strategies are excessively driven by available data without considering the previous

366   knowledge, leading to inefficient models. Other authors carried out external validation studies

367   but none of them made a critical appraisal (37–40). In a previous study, Varela et. al.

368   developed a prognostic model based on a systematic review of factors related to in-hospital

369   mortality. The model was built using a series of univariate meta-analyses that pooled adjusted

370   and unadjusted estimates altogether without taking into consideration the correlation among

371   these factors. These pooled univariate estimates were then transformed into risk points to

372   create a risk score (41,42). Our proposal includes more factors and our analysis included only

373   estimates from low risk of bias studies. All estimates are from multivariate adjusted models

374   and the weight each model has to build the meta-model is determined by their predictive

375   performance in a validation cohort. This statistical methodology is in concordance with

376   current recommendations (16,43).

377   *Implications for practice*

378   The decision whether to perform surgery for IE remains a challenge in clinical practice and it

379   should come after a careful balance between the procedural risk and its estimated benefit.

380   Critical preoperative state and priority of the procedure (urgent or emergency) are the most

381   salient risk factors included in our meta-model. Patients with depressed LVEF, NYHA, renal

382   failure have also worse prognosis. In addition, the aggressiveness of the IE infection as well

383   as the technical difficulties of the surgery also implied higher risk of mortality. We expect a

384   worse outcome in patients with IE caused by Staphylococcus or fungi or in patients with

385   paravalvular abscesses, fistulae or previous cardiac surgery because in these patients the

386 surgery is challenging. Although risk scores for predicting mortality do not offer help in terms

387 of establishing the burdens of surgical futility, they add a great value helping endocarditis

388 teams to manage this complex disease and lead toward more personalized assistance based on

389 individual patient characteristics. Moreover, the meta-model can be used to determine the

390 case-mix of surgical hospitals and compare their performance adjusted for their case-mix.

391 Although in the 2015 IE guidelines (44) the score created by De Feo-Cotrufo et al for native

392 IE is the only one recommended, it would be expected to change with the creation of several

393 new IE specific scores and the generation of a meta-model that outperformed existing models.

394 The interpretation of the meta-model coefficients should be interpreted with caution because

395 coefficients have been shrunk, related to "Stein's paradox" (45). Shrinkage introduces bias in

396 the multivariable regression coefficients, but if we shrunk properly ensure better predictions

397 (46)

398 *Challenges and opportunities*

399 Further external validation studies are necessary to confirm the improvement in predictive

400 ability of the meta-model. We will develop an online calculator to allow a simple and

401 effective use of the meta-model. Given the low incidence of infective endocarditis,

402 sufficiently large sample sizes for the adequate development of new predictive models are

403 difficult to come by. We encourage authors to make their data available in order to allow

404 building model based on available data (47).

405 **Conclusions**

406 The meta-model is a robust prognostic model to calculate the individualized risk of post-

407 operative mortality in patients with infective endocarditis. It was developed based on the

408 previous evidence using aggregation methods of the existing models identified from a

409 systematic review and after critical being appraised. The meta-model outperformed existing

410   models; therefore, this preoperative tool can help guide individually tailored choices made by

411   patients and clinicians.

Authors contributions

Conceptualization: BMFF, LVB, EGE, JLA, AM, JIP, AR, JZ; Search strategies: BMFF, NAD, JLA; Data extraction and Critical appraisal: BMFF, LVB, ACP; Methodology: BMFF, EGE, AM, JZ; Software, Formal analysis: BMFF; Validation: AM, JZ; Data adquisition/curation: BMFF, ENE, PM, MCF, MAG: Writing - Original draft: BMFF, EGE, JZ; Visualization: BMFF, LVB, NFH; Supervision: EGE, JZ; Writing – Review & Editing: All authors.

Bibliography

431

432   1.   Murdoch DR. Clinical Presentation, Etiology, and Outcome of Infective Endocarditis in
433   the 21st Century: The International Collaboration on Endocarditis–Prospective Cohort
434   Study. Arch Intern Med. 2009 Mar 9;169(5):463.

435   2.   Thuny F, Grisoli D, Collart F, Habib G, Raoult D. Management of infective endocarditis:
436   challenges and perspectives. The Lancet. 2012 Mar;379(9819):965–75.

437   3.   Iung B, Doco-Lecompte T, Chocron S, Strady C, Delahaye F, Le Moing V, et al. Cardiac
438   surgery during the acute phase of infective endocarditis: discrepancies between
439   European Society of Cardiology guidelines and practices. Eur Heart J. 2016 Mar
440   7;37(10):840–8.

441   4.   Chu VH, Park LP, Athan E, Delahaye F, Freiberger T, Lamas C, et al. Association between
442   surgical indications, operative risk, and clinical outcome in infective endocarditis: a
443   prospective study from the International Collaboration on Endocarditis. Circulation.
444   2015 Jan 13;131(2):131–40.

445   5.   Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to
446   systematic review and meta-analysis of prediction model performance. BMJ. 2017 Jan
447   5;i6460.

448   6.   Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al.
449   Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling
450   Studies: The CHARMS Checklist. PLoS Med. 2014 Oct 14;11(10):e1001744.

451   7.   Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting
452   Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med.
453   2009 Jul 21;6(7):e1000097.

454   8.   Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a
455   multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the
456   TRIPOD statement. BMJ. 2015 Jan 7;350:g7594.

457   9.   Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al.
458   Transparent Reporting of a multivariable prediction model for Individual Prognosis Or
459   Diagnosis (TRIPOD): Explanation and Elaboration. Annals of Internal Medicine. 2015 Jan
460   6;162(1):W1.

461   10.   Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search
462   Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance
463   Systematic Reviews. Smalheiser NR, editor. PLoS ONE. 2012 Feb 29;7(2):e32844.

464   11.   Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. (2020). EPPI-
465   Reviewer: advanced software for systematic reviews, maps and evidence synthesis.
466   EPPI-Centre Software. London: UCL Social Research Institute.

467 12. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST:
468   A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation
469   and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1.

470 13. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST:
471   A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann
472   Intern Med. 2019 Jan 1;170(1):51.

473 14. Muñoz P, Kestler M, De Alarcon A, Miro JM, Bermejo J, Rodríguez-Abella H, et al.
474   Current Epidemiology and Outcome of Infective Endocarditis: A Multicenter,
475   Prospective, Cohort Study. Medicine (Baltimore). 2015 Oct;94(43):e1816.

476 15. Breiman L. Stacked regressions. Mach Learn. 1996 Jul;24(1):49–64.

477 16. Debray TPA, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KGM. Meta-
478   analysis and aggregation of multiple published prediction models: Meta-analysis and
479   aggregation of multiple published prediction models. Statist Med. 2014 Jun
480   30;33(14):2341–62.

481 17. Martin GP, Mamas MA, Peek N, Buchan I, Sperrin M. A multiple-model generalisation of
482   updating clinical prediction models. Statistics in Medicine. 2018 Apr 15;37(8):1343–58.

483 18. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to
484   regression analysis based on fractional polynomials for modelling continuous variables.
485   Chichester, England ; Hoboken, NJ: John Wiley; 2008. 303 p. (Wiley series in probability
486   and statistics).

487 19. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues
488   and guidance for practice. Statist Med. 2011 Feb 20;30(4):377–99.

489 20. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.

490 21. Riley RD, Windt D van der, Croft P, Moons KGM. Prognosis research in healthcare:
491   concepts, methods and impact. 2019.

492 22. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development,
493   Validation, and Updating [Internet]. 2019 [cited 2020 Apr 28]. Available from:
494   https://doi.org/10.1007/978-3-030-16399-0

495 23. StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp
496   LLC.

497 24. De Feo M, Cotrufo M, Carozza A, De Santo LS, Amendolara F, Giordano S, et al. The
498   Need for a Specific Risk Prediction System in Native Valve Infective Endocarditis
499   Surgery. The Scientific World Journal. 2012;2012:1–8.

500 25. Gatti G, Benussi B, Gripshi F, Della Mattia A, Proclemer A, Cannatà A, et al. A risk factor
501   analysis for in-hospital mortality after surgery for infective endocarditis and a proposal
502   of a new predictive scoring system. Infection. 2017 Aug;45(4):413–23.

503 26. Madeira S, Rodrigues R, Tralhão A, Santos M, Almeida C, Marques M, et al. Assessment
504   of perioperative mortality risk in patients with infective endocarditis undergoing
505   cardiac surgery: performance of the EuroSCORE I and II logistic models. Interact
506   CardioVasc Thorac Surg. 2016 Feb;22(2):141–8.

507 27. Di Mauro M, Dato GMA, Barili F, Gelsomino S, Santè P, Corte AD, et al. A predictive
508   model for early mortality after surgical treatment of heart valve or prosthesis infective
509   endocarditis. The EndoSCORE. International Journal of Cardiology. 2017 Aug;241:97–
510   102.

511 28. Fernández-Hidalgo N, Ferreria-González I, Marsal JR, Ribera A, Aznar ML, de Alarcón A,
512   et al. A pragmatic approach for mortality prediction after surgery in infective
513   endocarditis: optimizing and refining EuroSCORE. Clinical Microbiology and Infection.
514   2018 Oct;24(10):1102.e7-1102.e15.

515 29. Olmos C, Vilacosta I, Habib G, Maroto L, Fernández C, López J, et al. Risk score for
516   cardiac surgery in active left-sided infective endocarditis. Heart. 2017
517   Sep;103(18):1435–42.

518 30. Gaca JG, Sheng S, Daneshmand MA, O'Brien S, Rankin JS, Brennan JM, et al. Outcomes
519   for endocarditis surgery in North America: A simplified risk scoring system. The Journal
520   of Thoracic and Cardiovascular Surgery. 2011 Jan;141(1):98-106.e2.

521 31. Martínez-Sellés M, Muñoz P, Arnáiz A, Moreno M, Gálvez J, Rodríguez-Roda J, et al.
522   Valve surgery in active infective endocarditis: A simple score to predict in-hospital
523   prognosis. International Journal of Cardiology. 2014 Jul;175(1):133–7.

524 32. Gatti G, Perrotti A, Obadia J, Duval X, Iung B, Alla F, et al. Simple Scoring System to
525   Predict In-Hospital Mortality After Surgery for Infective Endocarditis. JAHA [Internet].
526   2017 Jul [cited 2020 Dec 28];6(7). Available from:
527   https://www.ahajournals.org/doi/10.1161/JAHA.116.004806

528 33. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum sample
529   size for developing a multivariable prediction model: PART II - binary and time-to-event
530   outcomes. Statistics in Medicine. 2019 Mar 30;38(7):1276–96.

531 34. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the
532   sample size required for developing a clinical prediction model. BMJ. 2020 Mar
533   18;m441.

534 35. On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the
535   STRATOS initiative, Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg
536   EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019
537   Dec;17(1):230.

538 36. Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European
539   system for cardiac operative risk evaluation (EuroSCORE). European Journal of Cardio-
540   Thoracic Surgery. 1999 Jul;16(1):9–13.

541   37.   Varela L, López-Menéndez J, Redondo A, Fajardo ER, Miguelena J, Centella T, et al.
542   Mortality risk prediction in infective endocarditis surgery: reliability analysis of specific
543   scores†. European Journal of Cardio-Thoracic Surgery. 2018 May 1;53(5):1049–54.

544   38.   Pivatto Júnior F, Bellagamba CC de A, Pianca EG, Fernandes FS, Butzke M, Busato SB, et
545   al. Análise de Escores de Risco para Predição de Mortalidade em Pacientes Submetidos
546   à Cirurgia Cardíaca por Endocardite. ABC Cardiol [Internet]. 2020 [cited 2021 Mar 5];
547   Available from: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-
548   782X2020000300518

549   39.   Gatti G, Sponga S, Peghin M, Givone F, Ferrara V, Benussi B, et al. Risk scores and
550   surgery for infective endocarditis: in search of a good predictive score. Scandinavian
551   Cardiovascular Journal. 2019 May 4;53(3):117–24.

552   40.   Wang TKM, Oh T, Voss J, Gamble G, Kang N, Pemberton J. Comparison of contemporary
553   risk scores for predicting outcomes after surgery for active infective endocarditis. Heart
554   Vessels. 2015 Mar;30(2):227–34.

555   41.   Varela Barca L, Navas Elorza E, Fernández-Hidalgo N, Moya Mur JL, Muriel García A,
556   Fernández-Felix BM, et al. Prognostic factors of mortality after surgery in infective
557   endocarditis: systematic review and meta-analysis. Infection. 2019 Dec;47(6):879–95.

558   42.   Varela Barca L, Fernández-Felix BM, Navas Elorza E, Mestres CA, Muñoz P, Cuerpo-
559   Caballero G, et al. Prognostic assessment of valvular surgery in active infective
560   endocarditis: multicentric nationwide validation of a new score developed from a meta-
561   analysis. European Journal of Cardio-Thoracic Surgery. 2020 Apr 1;57(4):724–31.

562   43.   Debray TPA, Koffijberg H, Vergouwe Y, Moons KGM, Steyerberg EW. Aggregating
563   published prediction models with individual participant data: a comparison of different
564   approaches. Statistics in Medicine. 2012 Oct 15;31(23):2697–712.

565   44.   Habib G, Lancellotti P, Iung B. 2015 ESC Guidelines on the management of infective
566   endocarditis: a big step forward for an old disease. Heart. 2016 Jul 1;102(13):992–4.

567   45.   Efron B, Morris C. Stein's Paradox in Statistics. Scientific American. 1977;236(5):119–27.

568   46.   van Houwelingen JC. Shrinkage and Penalized Likelihood as Methods to Improve
569   Predictive Accuracy. Statistica Neerland. 2001 Mar;55(1):17–34.

570   47.   Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM, Cochrane IPD Meta-
571   analysis Methods group. Individual Participant Data (IPD) Meta-analyses of Diagnostic
572   and Prognostic Modeling Studies: Guidance on Their Use. PLoS Med. 2015 Oct
573   13;12(10):e1001886.

574